cDNA cloning and sequence analysis of hepatitis G virus genome isolated from a Chinese blood donor

WANG Hailin 王海林, CHEN Hongsong 陈红松, XIA Ningshao 夏宁邵, TAN Wenjie 谭文杰, CHEN Gang 陈 刚
LIU Yulan 刘玉兰, CONG Yu 丛 郁, SUN Jing 孙 婧, ZENG Ding 曾 定, HOU Yunde 侯云德
WANG Yu 王 宇 and ZHAN Meiyun 詹美云

Keywords: *HGVch* · *cloning* · *sequence analysis*

Objective To obtain full-length sequence of a Chinese hepatitis G virus (HGV) strain (HGVch) and investigate the genetic characteristic of HGVch and its identity to other isolates.

Methods Reverse transcription (RT) and nested-PCR were used to screen HGV RNA positive serum and amplify cDNA fragments. A positive serum without known hepatitis virus markers was selected for isolating HGV RNA template. The HGV genome was divided into 12 overlapping fragments and directly cloned into pGEM-T vector. Sequences were determined by dideoxy terminus-end method of DNA sequencing and then analyzed by computer.

Results The twelve fragments of HGVch cover 9213 nucleotides in length, containing a large open reading frame (ORF) encoding 2873 animo acids polyprotein that began with a methonine residue and ended at termination codon. HGVch is about 86.5%—89.5% identical to other known HGV isolates at the nucleotide level and about 93.9%—96.2% at the deduced animo acid level.

Conclusion HGV is a non-A-E hepatitis causal agent, proved to be related with posttransfusion hepatitis in all over the world. Chinese HGV isolate has very close relationship to other isolates from Africa, Europe, Japan, without significant difference across the entire genome. It is suggested that the sequences of HGV isolates are very conservative and the evolution is very slow.

Chin Med J 1999; 112 (8): 747-749

Although reliable assays are available for distinguishing hepatitis A-E virus, approximately 5%—10% of acute sporadic hepatitis cases and 20%—25% of chronic hepatitis cases in the US and 10%—20% of acute and/or chronic hepatitis cases in China are of unknown etiology,[1,2] suggesting that the existence of additional viral agents causing these cryptogenic hepatitis which have to be termed as non-A-E hepatitis. An RNA virus, designated hepatitis G virus (HGV),[3] was identified from a patient with chronic hepatitis in the early 1996; in the meantime, another RNA virus GBV-C[4] was isolated from a West African patient. Both of them were positive-stranded RNA virus with a genome approximately 9.4 kb in length, encoding a lengthy polyprotein of about 2800 animo acids; phylogenetic analysis demonstrated that HGV and GBV-C were additional members of the Flaviviridae, distinct from the HCV group. Homology between HGV and GBV-C were much higher than those of other flaviviruses in nucleotide and deduced animo acid sequence. So they were supposed to be two isolates of the same species of virus termed as HGV.

Recently, we reported that HGV RNA was detected in a number of serum sample of clinical patients and blood donors which were collected from different cities and countryside in China. The preliminary results of investigation suggested that HGV was another etiology agent of viral hepatitis with high prevalence in China. A complete sequence analysis of Chinese HGV genome is conducive to developing the specific and sensitive diagnosis methods of gene and serology, investigating molecular epidemiolgy as well as carrying out a

Institute of Virology, Beijing 100052, China (Wang HL, Tan WJ, Chen G, Cong Y, Hou YD and Zhan MY)

Institute of Hepatology, the People's Hospital, Beijing Medical University, Beijing 100044, China (Chen HS, Liu YL, Sun J and Wang Y)

Department of Biology, Xiamen University, Xiamen 361005, China (Xia NS and Zeng D)

This program was supported by the grants of the research project (102-07-02-07) of National High Technological Research and Development Program (863) and the Ninth Five-Year Plan.

Chinese Medical Journal 1999; 112(8): 747-749

basic research for HGV vaccine.

METHODS

Origin of serum sample

A serum sample was collected from a blood donor, a professional one for four years from Hebei province. Detected by RT-nPCR with a set of primers derived from NS3 helicase region of HGV genome, the serum was HGV RNA positive but without markers of other known hepatitis viruses. It was aliquated and stored at −70℃ before being used for RNA extract.

Genome extension

Primers which consisted of 12 sets or 45 oligonucleiotides were designed by OLIGO (Ver 5.0, NBI) referencing these nucleotide sequences (GenBank accession No: HGU36380, HGU44404). These primers were expected to get 12 cDNA fragments overlapping in flanking regions, which covered a complete genome. The positions of primers were defined according to HGU44404 (Fig. 1). HGV RNA was extracted from the serum and amplified by RT-PCR as previously described.[5,6]

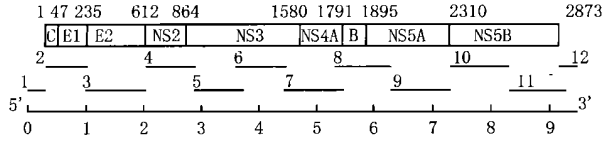

Fig 1. The position of cloning fragments of the HGVch genome and protein map of the deduced polyprotein.

DNA sequencing and sequence analysis

Three clones obtained from each cDNA fragment and extension experiments were sequenced with PE sequencing Kit (PE Co.) directed by manufacturer. The sequence was analyzed using programs in DNA, DNASIS and PROSIS software. The following genomic sequences were used: GBV-C(EA) (GenBank accession No: U36380), HGV FNP216 (U44402), HGV K10291 (U45966), GBV-C (EA) (U63715), HGVC964 (U75356), and HGV-J (U87255).

RESULTS

Analysis of nucleotide sequence

A total of 12 overlapping cDNA clones were isolated from the donor serum and sequenced. These cloned fragments covered 9213 nucleotides in length (GenBank accession No: 94695), including three parts: a large open reading frame encoding 2873 animo acids and two untranslation region (UTR) flanking large open reading frame, which are 423nt and 168nt at 5′ and 3′/NTR respectively. GC content of the sequence is 59.4%, much richer than AT content. Compared with other HGV genomes, HGVch exhibits 86.5% — 89.5% nucleiotide identity (Table), and most of this nucleotide variability within the coding region are silent and do not change conserved amino acids.

The 5′NTR of HGVch genome shows a high identity to the homologous region of HCV and GBV-A and B. The region is, however, regarded as a suitable region for genotype for its highest conservation (above 95%) in HCV genomes, but is slightly lower than the average homology among HGV genomes. These isolates may be divided into three distinct groups according to the percentage identity standard of 90% at 5′NTR (Fig. 2). The 3′NTR is little conserved in HCV genomes, but shows the highest conservation in HGV genomes, ranging between 89.8%—94.9%.

Table Nucleotide and deduced amino acid identity between HGVch and other HGV isolates

HGVch vs	GBV-C(WA)	GBV-C(EA)	PNF2161	R10291	HGVC946	HGV-J
5′NTR	87.5	90.6	88.7	88.5	89.5	89.0
C	90.6 (80.9) *	91.5 (74.5)	90.8 (89.4)	88.7 (80.9)	93.0 (87.2)	90.1 (89.4)
E1	87.1 (91.5)	86.5 (92.6)	83.7 (91.0)	83.9 (91.5)	91.9 (94.1)	82.5 (94.1)
E2	85.7 (91.5)	86.0 (94.2)	84.2 (89.1)	85.2 (92.6)	90.1 (94.7)	85.6 (93.1)
NS2	85.3 (93.6)	83.2 (94.0)	85.4 (93.7)	83.2 (95.2)	87.8 (92.9)	84.8 (93.7)
NS3	85.9 (94.7)	86.7 (97.9)	87.0 (97.5)	87.3 (97.6)	90.9 (96.8)	86.5 (97.8)
NS4	85.9 (94.6)	86.7 (96.5)	85.3 (95.3)	86.1 (96.8)	86.1 (97.1)	85.5 (96.2)
NS5	88.7 (95.3)	88.5 (97.4)	87.6 (96.8)	88.1 (96.9)	91.4 (96.2)	88.1 (97.5)
3′NTR	93.0	93.6	94.9	89.8	—	94.2
Total	87.0 (93.9)	87.4 (96.0)	86.5 (95.0)	86.7 (95.8)	89.5 (95.7)	86.7 (96.2)

* the number within brackets are the identitiy (%) of amino acid sequences.

Analysis of polyprotein

In HCV, polyprotein is presumed to be processed into the structural and nonstructural proteins by host and viral proteases. And this is also supposed to be the case for HGV because its genome organization is smiliar to HCV genome. Within the inferred structural region of HGVch, there are

four eukaryotic signal sequence cleavage sites in C/E1, E1/E2, E2/NS2, and within E2 for P7 (Fig. 1). The cleavage sites in nonstructural proteins are smiliar to that in HCV, except that host protease for E2/NS2, additional processing of the HCV nonstructural proteins (NS3-NS5B) occurs via a second viral serine protease, where the motif containing His (1083) and Ser (1165) is very conserved. Multiple alignment analysis confirmed the presence of the serine protease motif in NS3 of HGVch and other HGV isolates and several amino acid stretches that may serve as substrates for this putative enzyme (Fig. 1).

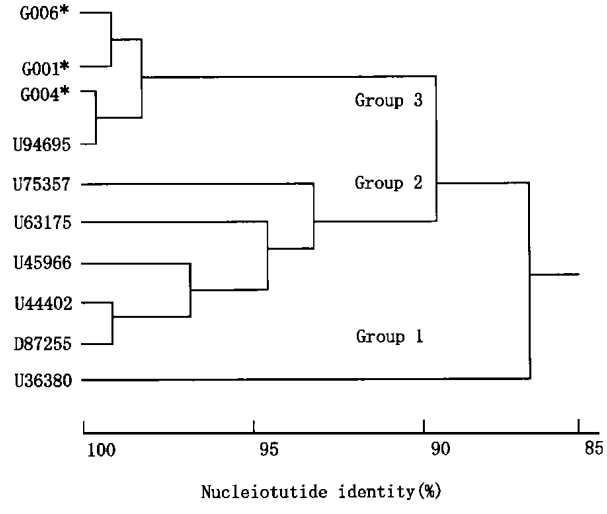

Fig 2. The predicted HGV genotypes based on identity of 5′NCR sequences. * sequences derived from reference 8.

A detailed analysis to identify overall and regional sequence variability between HGVch and other known isolates was performed by alignment (Table). The nonstructural protein including NS3, NS4 and NS5 proteins showed slightly higher identity than the mean identity of 97% approximately. Area of E1, E2, and NS2 proteins showed conservation slightly lower than the average. The C protein consisted only of 47 amino acids, much different from that of HCV. The putative E2 protein exhibited the same degree of diversity of about 8% as the E1 protein, and has no High Varibility Region (HVR). Besides glycosylation sites in E1 and E2 proteins, there were one site in NS3 and three in NS5, much like that in HCV.

DISCUSSION

Immediately after the genomes of two positive-strand RNA viruses, GBV-C and HGV (GenBank accession No: HGU36380 and HGU44404, respectively) were reported to be cloned and sequenced by two viral discovery groups separately in early 1996 as a new hepatitis virus. Our research group has proved the extensive prevalence of HGV in clinical hepatitis and even in blood donors in China.[6] So it is urgent to develop a serological and molecular diagnosis methods for antibodies anti-HGV and HGV RNA. First of all, the genome of Chinese HGV strain should be analyzed and then molecular epidemiology be investigated with specific and sensitive HGV diagnosis assay for gene and anti-HGV antibodies.

The HGVch we have cloned and sequenced stands for another Chinese HGV strain. The near full-length genome is 9213 nucleotides. To contrast HCV genomes, the most conserved region is 5′NTR and the 3′NTR only shares the lowest identity. However, the 5′NTR of 423 nucleotides of HGVch has about 90% identity to other HGV isolates, slightly lower than average percentage of homology, and the 3′NTR of 168 nucleotides has a higher identity. In contrary to the E2 of HCV genome, the E2 of HGV genome does not possess the so called HVR. The pupative precursor polyprotein of 2873 amino acids contains some protein motifs conserved in other positive-strain RNA viruses, such as serine protease, RdRp and cleavage sites. Interestingly, the core protein in HGV polyprotein is very short and only has 47 amino acids. The HGV may be lack of intact core protein encoded by itself and the capsid may be assembled by hijacking other viral or host membrane. However, it is possible that the short fragment is a part of capsid because antibody against the oligopeptide is detected in sera of HGV RNA positive hepatitis.

Although the HGV isolates were divided into three groups, but the genome of HGVch possesses high homology with the genomes isolated from geographically distinct regions. Such results suggest that HGV genome is very conservative and evolution is very slow.

REFERENCES

1. Decker R, Dawson G, Mushahwar I. GBV-C: a new human hepatitis-related virus. Today's Life Sci 1995; 21: 20.

2. Moaven L, Bowden DS, Locarnini SA. Hep G virus and hep GB viruses. Today's Life Sci 1995; 21: 24.

3. Linnen J, Wages JR, Zhang-Keck ZY, et al. Molecular cloning and disease association of hepatitis G virus: a transfusion-transmissible agent. Science 1996; 271: 505-508.

4. Leary TP, Muerhoff AS, Simmons JN, et al. Sequence and genomic organization of GBV-C: a novel member of the Flaviviidaes associated with human non-A-E hepatitis. J Med Virol 1996; 48: 60-67.

5. Erker CE, Simmons JN, Muerhoff AS, et al. Molecular cloning and characterization of a GB virus C isolate from a patient with non-A-E hepatitis. J Gen Virol 1996; 77: 2713-2720.

6. Wang HL, Xia NS, Tan WJ, et al. Detection of hepatitis GB virus C (GBV-C) RNA in Chinese hepatitis patients. Chin J Virol 1996; 12: 183-185.

(*Received March 15, 1998*)
本文编辑: 王 德